

The unreasonable effectiveness of near infrared spectroscopy explained. Introducing the era of NIRS integrated functional fingerprinting

Lars Munck

Spectroscopy and Chemometrics Group, Institute of Food Science, University of Copenhagen, Frederiksberg, Denmark. E-mail: Imu@food.ku.dk

Near infrared (NIR) spectral patterns from seeds carry surprisingly consistent information on functional food quality that seems almost unreasonable. In the developing cereal seed, all genes interact with all other active genes. This is proven by manipulating the immense complexity of physiological expression by single seed mutants in a barley seed model at a constant gene background/environment. Miraculously, it is possible to get a causal relation between the mutated gene and its NIR spectral pattern. When combining under controlled conditions, Nature's and the near infrared spectroscopy (NIRS) instrument's combined ability to reproduce the physiochemical composition of a seed population of a DNA-specific genotype, we obtain at stunning absorption log $1/R \ 10^{-4} - 10^{-5}$ mean distance between two similar single seed spectra. The deterministic gene/genotype-specific NIRS patterns from single seeds/populations and the corresponding deterministic metabolite patterns are interpreted, in the light of the present achievements in molecular and quantum biology, as a discovery of a global decision mechanism by "set probability". It involves a global cell/seed/plant communication resulting in individual specific NIRS and chemical metabolic patterns that prevail on all levels of the phenotype. This ensnaring phenomenon is here labelled as "biological entanglement" with a possible explanation in the quantum world of physics. The observed gene interaction leading to conserved physiochemical fingerprints has consequences with regard to the assessment and breeding for cereal food quality. Only by targeting the food functional complex by a spectral pattern from a certified quality cereal line, that includes an overview of gene interaction, is it possible to get the selection target right from the beginning. Selection by functional NIR spectral fingerprinting first is more effective than by quantitative trait locus (QTL) markers for single quantitative traits, because these DNA markers do not include information on the global gene interaction. The cost of using integrated NIRS breeding is less than one per cent of the costs for molecular QTL breeding. For a new NIRS fingerprinting culture to take on in industry, plant breeding and basic science-an inverse engineering strategy is needed with a library for functional fingerprint evaluation that builds a confidence into each user based on own experience, on how to read the meaning out of every single NIR spectrum.

Introduction

I enjoyed in the 1970's at the Carlsberg Research Centre in Copenhagen the great success of *the* 1st *NIRS culture*: *Analytical Chemistry*. Near infrared spectroscopy (NIRS) substituted for *the* crude analyses such as protein, starch, fiber and fat that were developed in the 19th century that now have a much more complex identity. However, their commercial importance is still fully retained, which gives a conservative edge to the NIRS applications developed from them.

Correspondence

L. Munck (<u>Imu@food.ku.dk</u>) doi: 10.1255/nir2017.105

Citation: L. Munck, "The unreasonable effectiveness of near infrared spectroscopy explained. Introducing the era of NIRS integrated functional fingerprinting", in *Proc.* 18th Int. Conf. Near Infrared Spectrosc., Ed by S.B. Engelsen, K.M. Sørensen and F. van den Berg. IM Publications Open, Chichester, pp. 105–114 (2019). https://doi.org/10.1255/nir2017.105 © 2019 The Author

The spectroscopic and chemical skills of Karl Norris and Phil

Williams¹ resulted in chemical assignments by second deriva-

tives that emphasize visual validation of spectra. The assign-

ments were exploited by multiple linear regression prediction

of analytes for quality monitoring. The breakthrough came

in 1975 when the Canadian Grain Commission adopted NIRS for hundred thousands of polluting laboratory anal-

yses for Kjeldahl protein. In 1976-1986 Harald Martens and



ISBN: 978-1-906715-27-4

This licence permits you to use, share, copy and redistribute the paper in any medium or any format provided that a full citation to the original paper is given, the use is not for commercial purposes and the paper is not changed in any way.

Svend Aage Jensen, then at Carlsberg, initiated the 2nd NIRS culture: Chemometrics by PLSR (partial least squares regression). The first paper² on the use of NIRS to predict Kjeldahl protein analysis was published in 1983, 100 years after Johan Kjeldahl's at the Carlsberg Laboratory. PLSR modelling was highly convenient and competitive because it allowed inexperienced users successfully to exploit NIRS for measurement of components in biological materials without considering assignments and without looking at NIR spectra. The disadvantage was that the art of visual-chemical validation by NIR spectra from the 1st NIR culture was lost.

The aim is here to visualize casual interpretable patterns in NIR spectra from seeds³⁻⁵ that mirror the underlying precision in the self-organized physiochemical composition, softly "calculated" by Nature,⁵ by introducing a pertubation mutant barley model.⁶ It is here necessary to reintroduce and further develop a science in spectral inspection from the 1st NIRS culture, but now, instead of assignment of chemical bonds to wavelengths, based on identification of functional factors as irreducible patterns in whole NIR spectra by conceptual validation³ that gives biological and technological meaning. We can then envisage the rise of a 3rd NIRS culture: Functional Fingerprinting. The motto relies on selection by spectral fingerprinting FIRST-validation of selected seed samples by physiochemical analyses and genetic molecular markers SECOND (if required).

Experiments and results

Material and methods

A barley mutant model visualizing gene/environment expression by NIRS. Since the 1960's about 30 "high lysine" barley seed mutants were selected⁶ from a Kjeldahl protein (KP) versus dye-binding (DBC) scatter plot. The mutant material including mutant breeding lines was collected at Carlsberg and kept at University of Copenhagen (KU) comprising 230 lines including normal **N** barleys. The mutants selected for lysine revealed surprisingly not only the regulative protein/very high lysine **P** mutants, but also structural carbohydrate **C** mutants.⁷ These are affecting phosphorylation and transport of ADP-glucose, resulting in significant decrease of starch and a compensatory increase in β -glucan.⁷ All **P** and **C** mutants have shriveled seeds with 80–70% in yield. The Carlsberg barley material⁶ from 1991 was multiplied in 1998–2004 at KU in two extreme conditions in pots in a greenhouse (3 years, outdoors 1 year) and in the field (3 years) to get maximal variation in chemical composition. A breeding material of *lys3.a* **P** recombinants including mutant recombinants from Carlsberg selected for plump seeds is included together with **N** normal controls.

Chemical analyses.⁸ Laboratory analyses⁷ were made on representative samples of the total material of about 520 samples and used for calibrating a PLSR model to predict water, protein, amide-N, starch, β -glucan, fat and rest fraction % d.m.

NIR spectral analyses.⁸ The seeds were equilibrated to the moisture of the laboratory at 23–24 °C. Thus water % d.m. represents a relative measurement on water activity that brought precision and interpretability to NIRS patterns. Finely milled seed samples were measured by a NIRSystems 6500 (Foss A/S, Hillerød, Denmark) spectrometer 400–2500 nm⁸ using a rotary cup with quartz window at 16 references and 32 samples scans.

Chemometrics.^{5,8} Local spectra were scatter-corrected by multiplicative scatter correction (MSC). Chemical analyses were predicted from calibrations models with laboratory analyses by PLSR.⁵ Classification was performed by principal component analysis (PCA)^{3,6,8} and by spectral inspection of differential spectra.

Integrated physiological expression of genes, environment and functional quality revealed by finely tuned chemical and NIR spectral patterns

Williams (p. 8)¹ recognized the importance of keeping in close contact with visual spectral validation and identified over 50 external facts from cereal samples that influenced the shape of the NIR spectra. He⁹ documented 24 of them in spectral plots over the entire region from 400 nm to 2500 nm. This format was also used in Williams and Norris classic handbook on Near Infrared Technology.¹ However, the aim of the NIRS handbook¹ by Williams and Norris was not to document genetic fingerprints, but to define wavelength assignments specific for chemical bonds and chemical components in analytical chemistry. Therefore, the log(1/R) spectra in the handbook supplements¹ include a corresponding second derivative spectrum inspired by Karl Norris. When five scatter-corrected spectra from the barley model are introduced at low magnification in Figure 1A the variation between the N, P and C mutant lines is hardly visible. This spectrum mirrors the typical barley spectrum A3 in the handbook (p. 284).¹

0.6

0.4

0.2

0

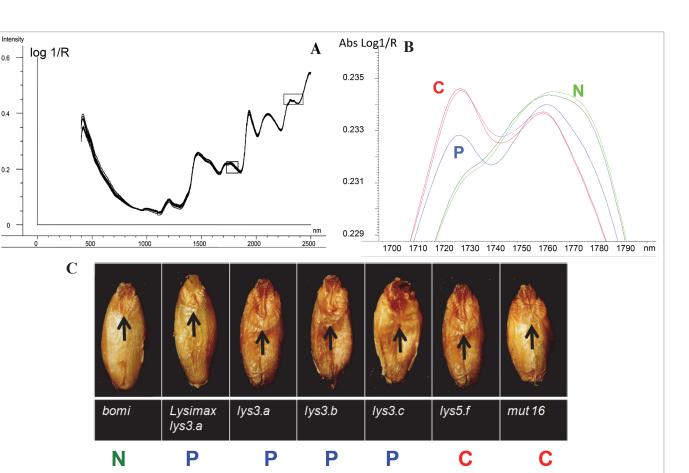


Figure 1. NIR reflection MSC scatter-corrected spectra 400–2500 nm for N P C genotypes. The "hot" information spots 1710-1790 nm (B) and 2260-2360 nm (Figures 2 and 4A) are marked. B Characteristic NIR spectral patterns MSC 1710-1790 nm for C structural carbohydrate barley mutants *lys5.f* (in red) and Risø mutant 16 (violet) and P regulatory protein mutant lys3.a (blue) all compared to spectra of normal N barleys, mutant parent Bomi (light green) and variety Triumph (dark green) grown together in pots in greenhouse. C Seed morphology of normal N barley Bomi and its shriveled mutants seeds P protein mutant alleles lys3.a, lys3.b and lys3.c and C carbohydrate mutants lys 5.f and Risø mutant 16. Lysimax is an improved seed recombinant from lys3.a x normal barley crosses selected for plump high starch seeds.

However, when we zoom in on a small spectral area 1690-1810nm indicated in Figure 1A and enlarge it 21 times in Figure 1B, we enter the world of the barley NIR spectral phenome⁷ by conceptual validation³ to assign biological meaning to the characteristic spectral patterns that are unique for each genotype and sample.

Information from high-resolution NIR spectral patterns have been underestimated.⁸ We can now at 21× magnification in Figure 1B recognize specific MSC log1/R spectral patterns for each of the N C P mutant classes. The spectral patterns of the carbohydrate mutant C spectra, lys5.f and mutant 16, clearly form a separate group compared to the normal N spectra Bomi and Triumph with the **P** protein mutant *lys3.a* mutant in between. The change in NIRS patterns between the three groups reflects a stunning change in chemical composition from

the normal **N** parent Bomi (starch 48.5%, β -glucan 6.8%, lysine/16gN 3.3%) to its C carbohydrate mutant lys5.f (starch 29.8%, β-glucan 19.8%, lysine 3.3/16gN %) and further to its deviating protein lysine **P** mutant lys3.a (starch 40.4%, β-glucan 4.7%, lysine/16gN 4.9%).

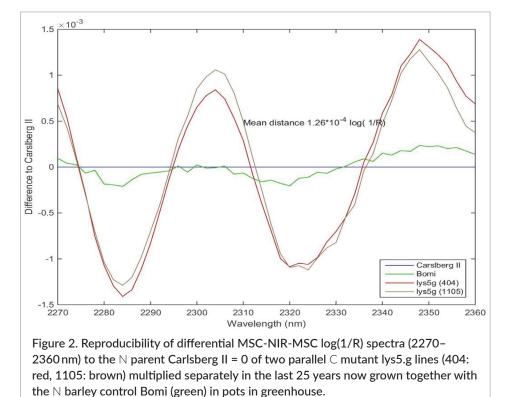
The morphological expression of seed mutations. The quite radical C and P mutants⁶ give visible effects on seed morphology (Figure 1C) with shrunken seeds and a large embryo scutellum plate (arrow) that leads to increased fat content and low yield (75% of normal). The morphological visual characters for the high lysine mutant genes are qualitatively recessively inherited 3:1 (normal to mutant). However, all genes interact more or less with thousands of quantitative genes during seed development that are not per se regularly inherited, leading to an integrated mutant syndrome in chemical composition

reflected in the specific patterns of NIR spectra (see above) and in electron microscope micrographs.⁶ Lysimax is an improved *lys3.a* recombinant variety from a plant breeding program at Carlsberg 1976–1991 selected for plumpness and yield.

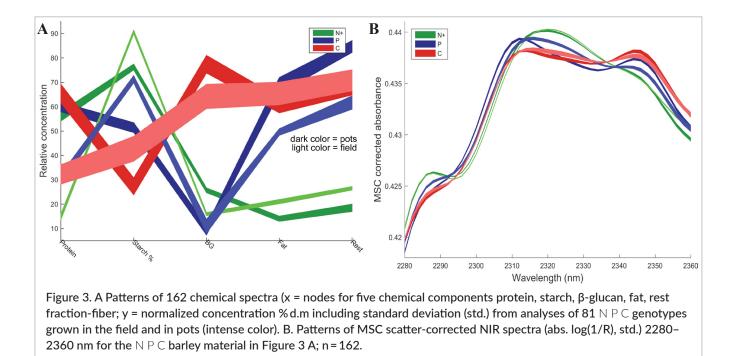
The integrated deterministic reproducibility of the whole seed NIR spectral measurement system

I will first focus on the unreasonable efficiency^{3,8} of our classic Foss Near Infrared 6500^{7,9} spectrograph to demonstrate the deterministic reproducibility of scattercorrected spectra from the ripe seed population samples shown in Figure 2. Two genetically identical C lys5.g lines 404 and 1105 has been multiplied separately in the last 25 years. They are in Figure 2 evaluated as grown in the same field together with normal N parent Carlsberg II and control Bomi. The figure displays differential NIR spectra 2270-2360 nm for the sub-lines 404 and 1105 with the NIR spectrum of the parent Carlsberg II set to zero, all grown together with the N control Bomi in the same field. We find a staggering deterministic mean reproducibility of the NIR spectra between the two lines by $1.3 \log(1/R)$ 10⁻⁴ absorption units, including instrument error. These spectra represent the reproducibility of the exorbitant complex composition of the barley organism in Nature

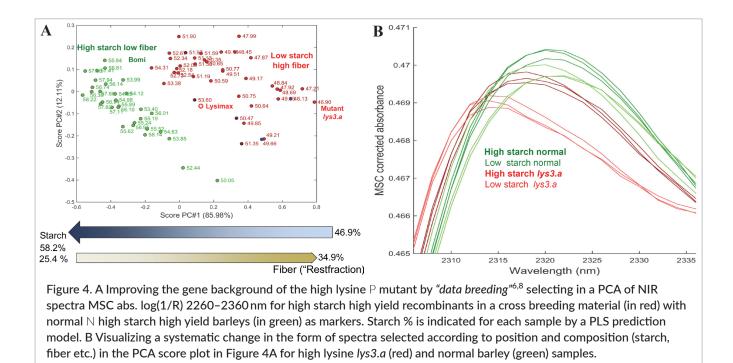
from seed to plant to seed amalgamated with the impressive instrumental reproducibility of the instrument.⁸ This example of the "unreasonable effectiveness of NIR spectroscopy" has been confirmed with NIRS analyses on barley material grown in the field and pot conditions for three and, respectively, four years. We have found that single seeds from a homozygous line grown in the same field have a conserved NIT (near infrared transmission) spectrum¹⁰ that characterizes the genotype of the biological unit (as in Figure 1B) in spite of a staggering variation in composition of single analytes, e.g. 6.3-17.9% for protein. This is surprising because the single seed variation in chemical composition in our experiment is just as high as that between the 162 seed populations of normal and mutant lines in Figure 3A. This means that the whole complex physiochemical gene expression process from sowing seed to plant population back to the ripe seeds are deterministically reproducible on the population level for each genotype in the same environment. Because DNA of the seeds in a population from a self-fertilizing homozygous barley line (variety) is nearly identical, the huge differences in single seed chemical composition are due to the effect of the timing of gene expression during seed synthesis (epigenesis) due to varying microclimate, sunlight and nutrition. This large variation in single seed physiochemical composition is the basis for the success



108



of NIT single seed sorting. Instead of looking on results as numbers in a table with standard deviations, we can now cognitively improve the evaluation of a seed population material including the original mutants and crosses of 162 ripe N C P lines by presenting the corresponding patterns as "chemical spectra" in Figure 3A. The material was grown in two contrasting environments field and pots. Here we are plotting at five nodes on the x axis the normalized chemical y values (range set to 100) for each of the 162 samples for protein (range 9.7-31.9%) starch (22.8-59.0%), β-glucan (2.2-17.8%), fat (1.6-4.2%) and rest (fiber 24.6-36.6%). We can conclude that in spite of the great compositional difference due genetics and environment (pots high and field low in spectral color intensity) the chemical spectra show that the genetic N C P message is dominating and well conserved. The thickness of the spectra represents the statistical standard deviation of each population. Here the variation in the N spectra for normal varieties is narrow due to a small genetic variation that is much broader in the C carbohydrate mutant spectra that are genetically more diverse. The environmental difference is greater for **P** mutants compared to C and N. Looking at the protein node for the chemical spectra with five nodes (Figure 3A) there is no tendency for a genetic trend that, however, gradually develops into significant N C P patterns when starch, β-glucan and fat come into play. Thus, the individuality of the chemical fingerprint increases with increasing information that defines the "multivariate advantage".³ These results end in a specific deterministic chemical pattern for each single individe that can be trusted, confirmed by the stunningly combined natural and instrumental reproducibility of the lys5.g NIR spectra in Figure 2. The scatter-corrected NIRS patterns in the 2280-2360 nm region in Figure 3B from the same barley material speak a corresponding conserved language as the chemical spectra, however, now much finely tuned with 40 information points (every second wavelength). The conclusions regarding the influence of N P C and environment on the patterns of spectra are the same in Figures 3A, B. We now understand that it is impractical to select for single chemical components or genes one by one at a time in quality control and breeding because the rigid gene interaction¹¹—to conserve homeostasis—transfers the selection for one component into a global chemical pattern characteristic of each genotype/environment combination. Thus, conventional selection in plant breeding¹¹ one by one, for physical/chemical analytes and for specific genes and quantitative trait locus (QTL)-DNA markers within a functional complex, e.g. for baking quality defined by 6-10 analyses, is not an optimal tool. This is because single genes and analytes lack information on gene expression interactions, but they are included in NIR spectra.¹¹ To get it right from the beginning, our experiment tells us that we have to use a NIR spectral pattern from a superior variety with certified quality in a genetically diverse material



to select for and find the specific recombinant having a gene combination that optimally combines all the parts of the functional complex. A specific NIR spectrum is now the conceptual label³ for food quality that can be used as a marker for selection. We should in the analytical culture since Norris and Williams, extend chemical wavelengths assignments with those of single NIR spectral patterns for functional criteria that should be trusted as much as the assignments.

From selection by molecular QTL-DNA markers for single quantitative physiochemical traits to integrated selection by functional NIR patterns in a PCA score or directly by visual inspection of scattercorrected NIR spectra

The *lys3.a* mutant gene has the most nutritious protein of all cereals, but it has severe pleiotropic side effects on the expression of other important genes that leads to shriveled seeds (Figure 1C), low yield and starch percentage and high fiber content. NIRS patterns can make integrated quality breeding possible to correct the gene background of the mutant by selecting for the global gene expression of a *lys3.a* breeding material in a PCA (see Figure 4A). In the PCA, selection for improved starch recombinants is demonstrated from crosses between the high lysine, shriveled, low starch **P** mutant *lys3.a* and

normal plump high starch N barley lines all grown in the field. The PCA classification is based on scatter-corrected (MSC) NIR spectra 2260-2360nm from 58 high lysine recombinants from crosses marked in red. Starch content for each sample is inserted from a PLS prediction model based on chemical analyses. Spectra from 30 normal N barleys in green are included as markers for high starch genotypes to increase the sensitivity of the PCA classification. The differences in chemical composition in the PCA (Figure 4A) involves the integrated whole genetic chemical expression of the genotypes as shown by the span of variation from right to left in the PCA for starch from 46.9% to 54.3% (with typical spectra shown in Figure 4B) and for fiber (rest fraction in Figure 3A) from 34.9% to 25.4%. It can be concluded that a PCA score plot is self-calibrating and the response depends on the diversity of the spectral patterns that are included. We can now identify the recombinant Lysimax in the middle of the PCA with an improvement in starch from 46.9% to 53.6%. Because high starch content indicate plumpness (Figure 1C) and increased yield, Lysimax was shown in field trials in 1989 to have a yield of +105% compared to the controls with a high lysine content of 5.4% g/16gN (normal barley 3.0-3.5%). Today the analytical costs of integrated NIRS data breeding that includes gene interaction (pleiotropy), e.g. for β -glucan in oats, is less that one percentage of the price for QTL-DNA marker breeding

that can not model gene interaction on the level of the physiochemical phenotype. When the 3rd NIRS fingerprinting culture is introduced, data applications will be developed to pay for functional quality from NIRS fingerprints from chemically specified high-quality genotypes as markers. Only by NIRS fingerprint breeding is it possible to include the important gene interaction effects in selection.¹¹ Now a new TriQ industrial single seed sorting equipment can verify the unreasonable efficiency of NIT spectroscopy by functional sorting of wheat with about 25,000 seeds per second (2.5 tons h^{-1}). The result from a pilot sorter for baking quality in wheat¹² is displayed in Table 1. The machine is self-calibrating, exploiting the extremes in chemical composition of the seed population to sort three fractions with one third of the input each. The result is impressive. The response range is 52.5-55.1% for water absorption from 14.8% to 25.0% for wet gluten and 1.9 to 2.5 in development time. In a Canadian example with sorting 5000 tons toxic Fusarium DON infested wheat, single seed NIT sorting increases value from 0 to 265,000 Canadian dollars.¹³

Table 1. Single seed NIT sorting by a Bomill AB pilot seed sorter in a seed population of winter wheat (Vinjett one location) by near infrared transmission (NIT) pattern spectroscopy by a pilot version of the TriQ single Seed NIT sorter (150 kg h⁻¹).

	Wheat Vinjett	Sor	ting fraction		
	Origin	F1	F2	F3	
Yield %	100	35.1	32.8	32.1	
	Gra	in			
Protein %	11.2	9.4	10.7	12.7	
	Flo	Jr			
Protein %	10.2	8.8	10.0	11.9	
Wet gluten %	19.5	14.8	18.6	25.0	
	Dou	gh			
Water abs %	53.6	52.5	53.6	55.1	
Dev. time	1.9	1.6	1.7	2.5	
Stability	2.5	1.7	2.4	8.0	
Softening	98	125	85	55	

Data from Tønning et al., Cereal Chem. 86(6), 710 (2009).¹²

Discussion: explaining the unreasonable efficiency of NIR spectroscopy in biology

The renowned physicist Eugene Wigner¹⁴ stated "The Unreasonable Effectiveness of Mathematics in Natural Sciences" is valid for the non-living inanimate world. Wigner comments his attitude of deduction when approaching the animate world when he writes, "The principal emphasis (of mathematics) is on the invention of concepts (theorems). Mathematics would soon run out of interesting theorems if these have to be formulated in terms of the concepts which already appears in the axioms". However, the latter sentence is the definition of the tautology of soft modelling that is the basis of self-organization of Life. Soft modelling occurs in the organism when gene/environment interaction that penetrates the physiological field, expresses itself as irreducible deterministic patterns that here are confirmed a posteori by induction and abduction³ stabilized by "the multivariate advantage". Chemometrics is more competent in modelling soft biological problems without hard theorems, compared to the mathematics of Wigner, because chemometrics introduces a soft provisional theorem as an overview of *biological data* by a lattice of principal x-y components (principal components) that can map and define the scores within each data set. The question is now: how can soft modelling in real life itself supply a base for a theorem/reference? The regularities in compositional chemical and NIRS fingerprints from ripe (Figures 2, 3A, B, 4B) and developing seeds^{5,15} indeed reflect a decision interface of the coordinated organism that can serve as provisional theorem/reference in real time. The key point is to advance back from Albert Einstein's deterministic standpoint in macro cosmos that "God does not play dice" to the fluctuating probabilistic/stochastic quantum reality in micro cosmos that has to be extremely fast and coordinative to support Life.¹⁶ The deterministic DNA sequence connected to the correspondingly deterministic physiochemical phenotypic pattern (as in the staggering reproducible NIR spectra of the mutant seeds in Figure 2) should then as a whole be a product of the creativity in the physiological/biochemical decision interface by "set probability".³⁻⁵ Nature's sensing of the physical decision interface equal to the dynamic composi-

tion of the whole cell (seed) could be looked upon as the theorem that Wigner missed. This motivates the introduction of the term "biological entanglement" that stands up for itself by the communication of specific ensnaring gene and genotype patterns throughout the physiological expression of the organism that is proven here. To keep each piece of animate matter together, there is a need for a new understanding on how the pattern information is communicated in the quantum physics of the cell that is revealed by the rise of quantum biology.¹⁶ In recent experiments, communication by physical entanglement is studied within a green photosynthetic plant cell. It involves a global quantum field covering the whole cell kept together in real time by information transfer through "Quantum Beat" fluctuations. It serves as a pacemaker transferring information of the state (spin/resonance) of pairs of electrons and photons immediately mediated without involving matter or energy by physical entanglement (the EPR, Einstein, Podolsky Rosen, principle). It could instantly establish a connection between separated photon and electron pairs with a common history in the whole cell. Future research will prove the connection between the physical EPR entanglement and the concept of biological entanglement in the micro and macro world defined here "as one integrated communication phenomenon that can be *read by NIRS*". We have confirmed¹⁵ in a separate seed development study with 247 metabolites from the N P C barley experimental model that unique N P C

metabolic patterns lies behind the highly deterministic patterns in chemical and near infrared spectra observed here. They may be decided by the whole physiochemical composition communicated by the quantum field of the whole individe (cell, seed) as a floating theorem/reference in real time. Its mindboggling regularity explains an essential part of "The Unreasonably Efficiency of NIRS" in biology. Charles A. Miller in his chapter on Chemical Principles of Near Infrared Spectroscopy (page 22 of Reference 1) acknowledges that NIRS theory is incomplete by saying "Vibrational Spectroscopy made complicated is NIR made possible". In analogy, herewith we can here suggest "Life made utterly complicated by Nature is pattern recognition made simplified". Wigner wrote,¹⁴ "It is, as Schrodinger has remarked, a miracle that in spite of the baffling complexity of the world, certain regularities in the events could be discovered". Thus, animate world

simplifies its message by communicating patterns that can be evaluated with almost deterministic efficiency by near infrared fingerprinting (if the instrument is well calibrated). For the new 3rd NIRS fingerprinting culture to take off, an inverse engineering strategy is needed with a library for functional fingerprint evaluation that builds a confidence into each user on their own experience, on how to read the meaning out of every single NIR spectrum.

Conclusion

- Self-organization and genetics in Nature speaks for itself by physiochemical expression patterns that can be read by NIRS.
- 2) NIRS patterns in seeds speak the language of gene expression and reproduce gene expression with a deterministic precision.
- 3) Spectral patterns from previously analyzed highquality cereal varieties are used as markers for complex technological function.
- With NIRS in plant breeding there is no need for any specific calibration or gene markers, only high-quality reference lines.
- 5) Spectral fingerprints and a self-calibrating PCA on data are the selection interfaces in "data breeding" and in single seed sorting.
- NIRS pattern overviews are needed to complement molecular gene expression analysis with a measure for gene interaction.
- 7) The NIRS FIRST principle substitutes the reductionist approach in analytical chemistry and genetics for an integrated NIRS functional pattern greatly improving economy in plant breeding¹¹ and in grain quality control.⁸

Acknowledgments

This paper could not have been written without keen support by Åsmund Rinnan, Birthe Møller Jespersen and Lars Nørgaard to my research and in establishing the illustrations. I am grateful to Professor Søren B. Englesen in making my research at the University of Copenhagen possible since 2001 and to Dr P.C. Williams for valuable advice and inspiring discussions on NIR spectroscopy since the 1970's.

References

- P.C. Williams and K. Norris (Eds), Near Infrared Technology in the Agricultural and Food Industries. American Association of Cereal Chemists, St Paul, Minnesota, USA (2001).
- H. Martens and S.A.A. Jensen, "Partial least squares regression: a new two-stage NIR calibration model", *Proc.* 7th World Cereal Bread Congress, Ed by J. Holas and J. Kratochivil. Elsevier, Amsterdam, p. 607 (1983).
- L. Munck, "Conceptual validation of self-organisation studied by spectroscopy in an endosperm gene model as a data-driven logistic strategy in chemometrics", *Chemometr. Intell. Lab. Systems* 84, 26–32 (2006). <u>https://doi.org/10.1016/j.chemolab.2006.04.018</u>
- L. Munck, "A new holistic exploratory approach to systems biology by near infrared spectroscopy evaluated by chemometrics. and data expression", *J. Chemometr.* 21, 404–426 (2007). <u>https://doi.org/10.1002/cem.1079</u>
- L. Munck, B. Møller Jespersen, Å. Rinnan, M. Engelsen, L. Nørgaard and S.B. Engelsen, "A physiochemical theory on the applicability of soft mathematical models—experimentally interpreted", *J. Chemometr.* 24, 481–495 (2010). <u>https://doi.</u> org/10.1002/cem.1278
- 6. L. Munck and B.M. Jespersen, "From discovery of high lysine barley endosperm mutants in the 1960–70s to new holistic spectral models of the phenome and of pleiotropy in 2008", in *Induced Plant Mutations in the Genomics Era*, Ed by Q.Y. Shu. Food and Agriculture Organization, Rome and International Atomic Agency, Vienna, pp. 419–422 (2009).
- L. Munck, B. Møller, S. Jacobsen and I. Søndergaard, "Near infrared spectra indicate specific mutant endosperm genes and reveal a new mechanism for substituting starch with (1→3, 1→4)-β-glucan in barley", J. Cereal Sci. 40, 213–222 (2004). <u>https://doi.org/10.1016/j.jcs.2004.07.006</u>
- **8.** B.M. Jespersen and L. Munck, "Cereals and cereal products", in *Infrared Spectroscopy for Food Quality*

Analysis and Control, Ed by D.-W. Sun. Academic Press, Elsevier Inc., USA, pp. 274–318 (2009). https://doi.org/10.1016/B978-0-12-374136-3.00011-0

- P.C. Williams, "Grains and seeds", in *Near Infrared* Spectroscopy in Food Science and Technology, Ed by Y. Ozaki, W.F. McClure and A.A. Christy. Wiley-Interscience, Hoboken, New Jersey, USA, pp. 165–217 (2006).
- **10.** Å. Rinnan, Department of Food Science, University of Copenhagen, Rolighedsvej 26, DK 1958 Frederiksberg C, personal communication (2017).
- 11. L. Munck, "Breeding for quality traits in cereals a revised outlook on old and new tools for integrated breeding", in *Cereals Handbook of Plant Breeding*, Ed by M.J. Carena. Springer Publishers, New York, pp. 333–366 (2009).
- E. Tønning, A.K. Thybo, L. Pedersen, L. Munck, Å. Hansen, F.A. Tøgersen, S.B. Engelsen and L. Nørgaard, "Bulk functionality diversification by unsupervised single-kernel near infrared (SKNIR) sorting in wheat", *Cereal Chem.* 86(6), 706–713 (2009). <u>https://doi.org/10.1094/CCHEM-86-6-0706</u>
- 13. L. Munck, "Food security: Efficient removal of Fusarium-infested kernels by industrial single kernel sorting", *5th MONIQA International Conference 2015*, 16–18 September 2015, Porto, Portugal Food and Health Risks and Benefits. Internet address: <u>http://bomill.com/wp-content/</u>uploads/2016/08/20150916-ICC-Porto.pdf (2015).
- 14. E. Wigner, "The unreasonable effectiveness of mathematics in the natural sciences", *Commun. Pure Appl. Math.* 13, 1–14 (1960). <u>https://doi.org/10.1002/</u> <u>cpa.3160130102</u>
- 15. B. Khakimov, M.A. Rasmussen, R.M. Kannagara, B.M. Jespersen, L. Munck and S.B. Engelsen, "From metabolome to phenotype: GC-MS metabolomics of developing mutant barley seeds reveals effects of growth, temperature and genotype", *Scientific Reports* 7, 8195 (2017). <u>https://doi.org/10.1038/</u> s41598-017-08129-0
- **16.** J. Al-Khalili and J. McFadden, *Life on the Edge*—The *Coming Age of Quantum Biology*. Bantam Press, London (2014).